[19]中华人民共和国国家知识产权局

[51]Int. Cl[7]
G10L 15/00
G10L 13/00

[12]发明专利申请公开说明书

[21]申请号 01116524.3

[43]公开日 2002年11月13日

[11]公开号 CN 1379392A

[22]申请日 2001.4.11 [21]申请号 01116524.3

[71]申请人 国际商业机器公司
地址 美国纽约
[72]发明人 唐道南 沈丽琴 施 勤 张 维

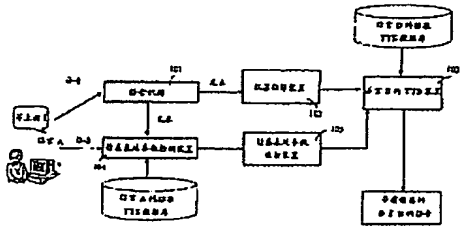
[74]专利代理机构 中国国际贸易促进委员会专利商标事务所
代理人 于 静

权利要求书3页 说明书11页 附图9页

[54]发明名称 具有情感的语音－语音翻译系统和方法

[57]摘要

本发明公开了一种具有情感的语音－语音翻译系统和方法。本发明的系统包括:语音识别装置、机器翻译装置、文本－语音生成装置、情感表述参数检测装置以及情感表述参数映射装置,其中,情感表述参数检测装置用于从原始语音信号中提取情感表述参数,而情感表述参数映射装置,用于将所述情感表述参数从一种语言(方言)映射到另一种语言(方言),并将映射结果作用于文本－语音生成装置,使其产生可以传达情感的语音输出。



ISSN1008-4274

1.一种具有情感的语音－语音翻译系统，包括：

语音识别装置，用于对语言 A 的语音表示进行识别，形成语言 A 的文本表示；

机器翻译装置，用于将语言 A 的文本表示翻译成语言 B 的文本表示；

文本－语音生成装置，用于根据语言 B 的文本表示生成语言 B 的语音表示，

所述具有情感的语音－语音翻译系统的特征在于还包括：

情感表述参数检测装置，用于从语言 A 的语音表示中提取情感表述参数；以及

情感表述参数映射装置，用于将情感表述参数检测装置提取的情感表述参数从语言 A 映射到语言 B，并将映射结果作用于文本－语音生成装置，使其产生可以传达情感的语音输出。

2.根据权利要求 1 的系统，其特征在于所述情感表述参数检测装置从不同层次提取情感表述参数。

3.根据权利要求 2 的系统，其特征在于所述情感表述参数检测装置从字、词级提取情感表述参数。

4.根据权利要求 2 的系统，其特征在于所述情感表述参数检测装置从语句级提取情感表述参数。

5.根据权利要求 1 的系统，其特征在于所述情感表述参数映射装置把所述情感表述参数从语言 A 映射到语言 B，然后再通过字词级变换映射和语句级变换映射将用于语言 B 的情感表述参数变换成用于调节文本－语音生成装置的参数。

6.一种具有情感的语音－语音翻译系统，包括：

语音识别装置，用于对一种方言 A 的语音进行识别，形成其文本表示；

文本－语音生成装置，根据所述文本表示生成另一种方言 B 的语音

表示；

所述具有情感的语音－语音翻译系统的特征还在于包括：

情感表述参数检测装置，用于从方言 A 的语音表示中提取情感表述参数；以及

情感表述参数映射装置，用于将情感表述参数检测装置提取的情感表述参数从方言 A 映射到方言 B 并将映射结果作用于文本－语音生成装置，使其产生可以传达情感的语音输出。

7.根据权利要求 6 的系统，其特征在于所述情感表述参数检测装置从不同层次提取情感表述参数。

8.根据权利要求 7 的系统，其特征在于所述情感表述参数检测装置从字、词级提取情感表述参数。

9.根据权利要求 7 的系统，其特征在于所述情感表述参数检测装置从语句级提取情感表述参数。

10.根据权利要求 6 的系统，其特征在于所述情感表述参数映射装置把所述情感表述参数从方言 A 映射到方言 B，然后再通过字词级变换映射和语句级变换映射将用于方言 B 的情感表述参数变换成用于调节 TTS 的参数。

11.一种能够传达情感的语音－语音翻译方法，包括以下步骤：

对语言 A 的语音表示进行识别，形成语言 A 的文本表示；

将语言 A 的文本表示翻译成语言 B 的文本表示；

根据语言 B 的文本表示生成语言 B 的语音表示，

所述能够传达情感的语音－语音翻译方法的特征在于还包括以下步骤：

从语言 A 的语音表示中提取情感表述参数；以及

将情感表述参数检测步骤提取的情感表述参数从语言 A 映射到语言 B，并将映射结果作用于文本－语音生成步骤，从而产生可以传达情感的语音输出。

12.根据权利要求 11 的方法，其特征在于所述情感表述参数检测步骤从不同层次提取情感表述参数。

13.根据权利要求 12 的方法，其特征在于所述情感表述参数检测步骤从字、词级提取情感表述参数。

14.根据权利要求 12 的方法，其特征在于所述情感表述参数检测步骤从语句级提取情感表述参数。

15.根据权利要求 11 的方法，其特征在于所述情感表述参数映射步骤把所述情感表述参数从语言 A 映射到语言 B，然后再通过字词级变换映射和语句级变换映射将用于语言 B 的情感表述参数变换成用于调节 TTS 的参数。

16.一种传达表征情感的语音－语音翻译方法，包括以下步骤:

对一种方言 A 的语音进行识别，形成其文本表示;

根据所述文本表示生成另一种方言 B 的语音表示;

所述能够传达情感的语音－语音翻译方法的特征还在于包括以下步骤:

从方言 A 的语音表示中提取情感表述参数；以及

将情感表述参数检测步骤提取的情感表述参数从方言 A 映射到方言 B 并将映射结果作用于文本－语音生成步骤，从而产生可以传达情感的语音输出。

17.根据权利要求 16 的方法，其特征在于所述情感表述参数检测步骤从不同层次提取情感表述参数。

18.根据权利要求 17 的方法，其特征在于所述情感表述参数检测步骤从字、词级提取情感表述参数。

19.根据权利要求 17 的方法，其特征在于所述情感表述参数检测步骤从语句级提取情感表述参数。

20.根据权利要求 16 的方法，其特征在于所述情感表述参数映射步骤把所述情感表述参数从方言 A 映射到方言 B，然后再通过字词级变换映射和语句级变换映射将用于方言 B 的情感表述参数变换成用于调节 TTS 的参数。

具有情感的语音－语音翻译系统和方法

本发明一般涉及机器翻译，具体地说涉及具有情感的语音－语音翻译系统和方法。

机器翻译是利用计算机使一种语言的文字或语音翻译为另一种语言的文字或语音的技术。即在语言学的关于语言形式和结构分析的理论基础上，依靠数学方法建立词典、语法并利用计算机巨大的存储容量和数据处理能力，在没有人工干预的情况下实现从一种语言到另一种语言的自动翻译。

目前的机器翻译系统通常是基于文本的翻译系统，即用于将一种语言文字翻译为另一种语言文字。但随着社会发展，需要基于的语音翻译系统，即能够进行语音－语音的翻译。可以利用现有的语音识别技术、基于文本的翻译技术以及 TTS（文本－语音）技术实现语音－语音的翻译，即，首先利用语音识别技术对第一种语言的语音进行识别，形成第一种语言的文本表示；使用现有的翻译技术将第一种语言的文本表示翻译成第二种语言的文本表示；再利用成熟的 TTS 技术根据第二种语言的文本表示产生第二种语言的语音输出。

然而，现有的 TTS（文本－语音）系统通常产生缺乏情感的单调的语音。在现有的 TTS 系统中，首先对所有字/词的标准发音按合成音记录并对此进行分析，然后在字/词级将用于标准“表述”的相关参数存储在字典中。通过字典中定义的标准控制参数和常用的平滑技术由各个合成分量产生合成的字/词。这种语音生成方式不能基于语句的含义和讲话者的情绪状态生成可以生动地表征情感的语音。

为此，本发明提出了一种具有情感的语音－语音翻译系统和方法。

根据本发明的具有情感的语音－语音翻译系统和方法，利用从原始语音信号中获得的情感表述参数驱动标准 TTS 系统，产生可以带有情感的语音输出。

本发明的一个目标是提供一种具有情感的语音－语音翻译系统，包括：语音识别装置，用于对语言 A 的语音表示进行识别，形成语言 A 的文本表示；机器翻译装置，用于将语言 A 的文本表示翻译成语言 B 的文本表示；文本－语音生成装置，用于根据语言 B 的文本表示生成语言 B 的语音表示，所述具有情感的语音－语音翻译系统的特征在于还包括：情感表述参数检测装置，用于从语言 A 的语音表示中提取情感表述参数；以及情感表述参数映射装置，用于将情感表述参数检测装置提取的情感表述参数从语言 A 映射到语言 B，并将映射结果作用于文本－语音生成装置，使其产生可以传达情感的语音输出。

本发明的再一个目标是提供一种可以传达情感的语音－语音翻译方法，包括以下步骤：对语言 A 的语音表示进行识别，形成语言 A 的文本表示；将语言 A 的文本表示翻译成语言 B 的文本表示；根据语言 B 的文本表示生成语言 B 的语音表示，所述能够传达情感的语音－语音翻译方法的特征在于还包括以下步骤：从语言 A 的语音表示中提取情感表述参数；以及将在情感表述参数检测步骤提取的情感表述参数从语言 A 映射到语言 B，并将映射结果作用于文本－语音生成步骤，从而产生可以传达情感的语音输出。

此外，本发明还提供了可以在同种语言的不同方言之间进行语音－语音翻译的方法和系统。

所述具有情感的语音－语音翻译系统包括：语音识别装置，用于对一种方言 A 的语音进行识别，形成其文本表示；文本－语音生成装置，根据所述文本表示生成另一种方言 B 的语音表示；所述具有情感的语音－语音翻译系统的特征还在于包括：情感表述参数检测装置，用于从方言 A 的语音表示中提取情感表述参数；以及情感表述参数映射装置，用于将情感表述参数检测装置提取的情感表述参数从方言 A 映射到方言 B 并将映射结果作用于文本－语音生成装置，使其产生可以传达情感的语音输出。

所述能够传达情感的语音－语音翻译方法包括以下步骤：对一种方言 A 的语音进行识别，形成其文本表示；根据所述文本表示生成另一

种方言 B 的语音表示；所述能够传达情感的语音－语音翻译方法的特征还在于包括以下步骤：从方言 A 的语音表示中提取情感表述参数；以及将情感表述参数检测步骤提取的情感表述参数从方言 A 映射到方言 B 并将映射结果作用于文本－语音生成步骤，从而产生可以传达情感的语音输出。

本发明的具有情感的语音－语音翻译系统和方法可以改善翻译系统或 TTS 系统的语音输出质量.

通过以下结合附图的说明，本发明的其它目标和优点将会更加清楚。详细的描述和具体的实施例只是为了进行说明而提供的，因为在本发明的精神范围内对于这些实施例的添加和改进对于本领域技术人员来说是显而易见的。

图 1 是根据本发明一优选实施例的具有情感的语音－语音翻译系统的方框图；

图 2 是根据本发明一优选实施例的图 1 中的情感表述参数检测装置的方框图；

图 3 是根据本发明一优选实施例的图 1 中的情感表述参数映射装置的方框图；

图 4 是根据本发明另一优选实施例的具有情感的语音－语音翻译系统的方框图；

图 5 是一流程图，描述了根据本发明一优选实施例的可以传达情感的语音－语音翻译过程；

图 6 是一流程图，描述了根据本发明一优选实施例的情感表述参数检测过程；

图 7 是一流程图，描述了根据本发明一优选实施例的情感表述参数映射以及调节 TTS 参数的形成过程；以及

图 8 是一流程图，描述了根据本发明另一优选实施例的可以传达情感的语音－语音翻译过程。

如图 1 所示，根据本发明一优选实施例的具有情感的语音－语音翻译系统包括：语音识别装置 101、机器翻译装置 102、文本－语音生成

装置 103、情感表述参数检测装置 104 以及情感表述参数映射装置 105. 其中，语音识别装置 101 用于对语言 A 的语音表示进行识别，形成语言 A 的文本表示；机器翻译装置 102 用于将语言 A 的本文表示翻译成语言 B 的文本表示；文本－语音生成装置 103 用于根据语言 B 的文本表示生成语言 B 的语音表示；情感表示参数检测装置 104 用于从语言 A 的语音表示中提取情感表述参数；并且，情感表述参数映象装置 105 用于将情感表述参数检测装置提取的情感表述参数从语言 A 映射到语言 B，并将映射结果作用于文本－语音生成装置，使其产生可以传达情感的语音输出。

正如本领域技术人员所熟知的，语音识别装置、机器翻译装置以及 TTS 装置都是可以使用现有技术来实现的。因此，以下只结合图 2 和图 3 描述一下根据本发明优选实施例的情感表述参数检测装置和情感表述参数映射装置。

首先介绍一下可以反映语音情感的关键性参数。可以在不同层次上定义反映语音情感的关键性参数。

1.在字/词级，反映语音情感的关键性参数有：速度（持续时间）、响度（能量级）以及基频（包括范围和音调）。注意，由于一个词通常由几个语音合成单元（在汉语中大多数词由两个以上字/音节组成），所以还必须在语音合成单元级以向量或时间序列的形式定义语音的情感表述参数。例如，当人们很生气时，他/她所说的字/词的响度就非常高，字/词的基频也比通常高，并且其包络不平滑，而且许多基频消失，同时持续时间变短。另一例子是，当人们在正常情况下说话时，可能会强调语句中的一些字/词，这样这些字/词的基频、响度、持续时间就会发生变化。

2.在语句级，我们将焦点放在语调上。例如，疑问句的包络不同于陈述句。

以下就结合图 2 和图 3 描述一下根据本发明一优选实施例的情感表述参数检测装置以及情感表述参数映射装置是如何工作的。即如何提取情感表述参数以及如何利用提取的情感表述参数驱动现有的 TTS 装置

产生能够传达情感的语音输出。

如图 2 所示，本发明情感表述参数检测装置包括以下模块:

模块 A: 分析说话者语音的基频、持续时间和响度。在模块 A，我们利用语音识别的结果进行语音和字/词（或字符）之间的对准。并按如下结构记录对准结果:

```
句子内容
{
    字/词编号
    字/词内容
    {  文本;
       文本的语音;
       字/词位置;
       字/词属性;
       语音开始时间;
       语音结束时间;
       *语音的波形;
       语音参数内容;
          {* 绝对参数;
             *相对参数;
          }
       }
    }
```

然后我们使用 Short Time Analyze(短时分析)方法得到如下参数:

1.每个短时窗口的短时能量。

2.检测字/词的基频包络。

3.字/词的持续时间。

由以上参数进一步得出:

1.字/词中平均短时能量。

2.字/词中最大的 N 个短时能量。

3.基频范围、最大基频、最小基频以及一个字/词中的基频数。

4.字/词的持续时间。

模块 B: 该模块根据语音识别的结果（文本），使用标准语言 A 的 TTS 系统产生不表征情感的语言 A 的语音。然后分析无情感 TTS 的参

数。以此参数作为基准。

模块 C: 分析有情感语音和标准语音之间以上参数的变化。其原因是不同人讲话的响度、基频以及速度可能不同，即使相同的人，在不同时间说相同的语句其参数也可能不同，所以在根据基准语音分析字/词在语句中的作用时，我们使用相对参数。

我们使用对参数进行归一化的方法从绝对参数中得到相对参数:

1.字/词中相对平均短时能量。

2.字/词中最大的 N 个相对短时能量。

3.字/词中相对基频范围、相对最大基频、相对最小基频。

4.字/词的相对持续时间。

模块 D: 根据来自标准语音参数的基准，在字/词级和语句级分析表述情感的参数。

1.在字/词级，我们比较有情感语音和标准语音之间的相对参数，以检测出哪些字/词的参数发生了大的变化。

2.在语句级，根据变化的等级以及字/词的特性对字/词排序，找出语句中关键的带有情感表述的字/词。

模块 E: 根据参数比较的结果和有关什么样的情感将引起哪参数变化的知识，得出句子的表征情感的参数，即检测出情感表述参数，并按以下结构记录:

```
情感表述信息
{
    语句的情感表述类型;
    字/词内容
     {  文本;
        情感表述类型;
        情感表述级;
        *情感表述参数;
     };
}
```

例如，当用汉语生气地说“闭嘴!”时，很多基频消失，并且其绝对响度大于基准，同时相对响度非常尖锐，持续时间大大短于基准，于

是可以在语句级得出该句子的情感为生气。情感表述关键词是“闭嘴”。

下面再结合图 3A，3B 描述一下根据本发明一优选实施例的情感表述参数映射装置是如何构成的。其包括：

模块 A：用于根据机器翻译的结果把表征情感的参数结构从语言 A 映射到语言 B。其关键是找出语言 A 中对于表述情感来说是关键的字/词对应于语言 B 中的哪些字/词。其映射结果如下：

```
语言 B 的语句内容
{
    语句情感表述类型;
    语言 B 的字/词内容;
    {   文本;
        文本的语音;
        在语句中的位置;
        在语言 A 中的字/词情感表述信息;
        在语言 B 中的字/词情感表述信息;
    }
}
语言 A 的字/词情感表述
{    文本;
    情感表述类型;
    情感表述级;
    *情感表述参数;
}
语言 B 的字/词情感表述
{
    情感表述类型;
    情感表述级;
    *情感表述参数;
}
```

模块 B：根据映射结果产生可以驱动语言 B 的 TTS 的调节参数，在此，我们使用语言 B 的情感表述参数表，其根据情感表述参数给出字/词的合成参数。表中参数是一相对调节参数。

具体过程如图 3B 所示，语言 B 的情感表述参数经过两级变换表（字/词级变换表和语句级变换表）变换之后形成用于调节 TTS 的参数。

两级变换表分别是：

1.字/词级变换表，用于将情感表述参数变换成调节 TTS 的参数，表的结构如下：

```
字/词 TTS 调节参数的结构
{
    情感表述参数类型;
    情感表述参数;
    TTS 调节参数;
};
TTS 调节参数的结构
{
    float Fsen_P_rate;
    float Fsen_am_rate;
    float Fph_t_rate;
    struct Equation Expressive_equat;(用于改变基频包络的曲线特性)
};
```

2.语句级变换表，用于根据语句的类型给出语句级上的韵律参数，该韵律参数可用于对上述字/词 TTS 调节参数做进一步调整。

```
语句级 TTS 调节参数的结构
{
    情感类型;
    字/词位置;
    字/词属性;
    TTS 调节参数;
};
TTS 调节参数的结构
{
    float Fsen_P_rate;
    float Fsen_am_rate;
    float Fph_t_rate;
    struct Equation Expressive_equat;(用于改变基频包络的曲线特性)
};
```

以上结合具体实施例描述了根据本发明的语音－语音翻译系统。正如本领域一般技术人员所认识别的，本发明还可以用于在同一种语言的不同方言之间进行语音－语音的翻译。如图 4 所示，该系统类似于图 1 所示的翻译系统，区别仅在于，在同种语言不同方言之间进行语音翻译

就不再需要机器翻译装置。具体地说，语音识别装置 101 用于对一种方言 A 的语音进行识别，形成其文本表示；文本－语音生成装置 103 根据所述文本表示生成另一种方言 B 的语音表示；情感表述参数检测装置 104 用于从方言 A 的语音表示中提取情感表述参数；并且，情感表述参数映射装置105用于将情感表述参数检测装置104提取的情感表述参数从方言 A 映射到方言 B，并将映射结果作用于文本－语音生成装置，使其产生可以传达情感的语音输出。

以上结合图 1－图 4 介绍了根据本发明的具有情感的语音－语音翻译系统，其利用从原始语音信号等中获得的情感表述参数驱动标准 TTS 系统，产生可以传达情感的语音输出。

本发明还提供了一种可以传达情感的语音－语音翻译方法。下面就结合图 5－图 8 描述一下根据本发明一个具体实施例的可以传达情感的语音－语音翻译过程。

如图 5 所示，根据本发明一优选实施例的可以传达情感的语音－语音翻译方法包括以下步骤：对语言 A 的语音表示进行识别，形成语言 A 的文本表示（501）；将语言 A 的本文表示翻译成语言 B 的文本表示（502）；根据语言 B 的文本表示生成语言 B 的语音表示（503）；从语言 A 的语音表示中提取情感表述参数（504）；以及，将情感表述参数检测步骤提取的情感表述参数从语言 A 映射到语言 B，并将映射结果作用于文本－语音生成步骤，从而产生可以传达情感的语音输出(505)。

以下就结合图 6 和图 7 描述一下根据本发明一优选实施例的情感表述参数检测过程以及情感表述参数映射过程。即如何提取情感表述参数以及如何利用提取的情感表述参数驱动现有的 TTS 过程产生可以传达情感的语音输出。

如图 6 所示，本发明情感表述参数检测过程包括以下步骤：

步骤 601：分析说话者语音的基频、持续时间和响度。在步骤 601，我们利用语音识别的结果进行语音和字/词（或字符）之间的对准。然后我们使用 Short Time Analyze（短时分析）方法得到如下参数：

1.每个短时窗口的短时能量。

2.检测字/词的基基频的包络。

3.字/词的持续时间。

由以上参数进一步得出:

1.字/词中平均短时能量。

2.字/词中最大的 N 个短时能量。

3.基频范围、最大基频、最小基频以及一个字/词中的基频数。

4.字/词的持续时间。

步骤 602: 根据语音识别的结果(文本),使用标准语言 A 的 TTS 过程产生不表征情感的语言 A 的语音。然后分析无情感 TTS 的参数。以此参数作为基准。

步骤 603: 分析有情感语音和标准语音之间以上参数的变化。其原因是不同人讲话的响度、基频以及速度可能不同,即使相同的人,在不同时间说相同的语句其参数也可能不同,所以在根据基准语音分析字/词在语句中的作用时,我们使用相对参数。

我们使用对参数进行归一化的方法从绝对参数中得到相对参数:

1.字/词中相对平均短时能量。

2.字/词中最大的 N 个相对短时能量。

3.字/词中相对基频范围、相对最大基频、相对最小基频。

4.字/词的相对持续时间。

步骤 604: 根据来自标准语音参数的基准,在字/词级和语句级分析表述情感的参数。

1.在字/词级,我们比较有情感语音和标准语音之间的相对参数,以检测出哪些字/词的参数发生了大的变化。

2.在语句级,根据变化的等级以及字/词的特性对字/词排序,找出语句中关键的带有情感表述的字/词。

步骤 605: 根据参数比较的结果和有关什么样的情感将引起哪参数变化的知识,得出句子的表征情感的参数,即检测出情感表述参数。

下面再结合图 7 描述一下根据本发明一优选实施例的情感表述参数映射过程。其包括:

步骤 701：用于根据机器翻译的结果把表征情感的参数结构从语言 A 映射到语言 B。其关键是找出语言 A 中对于表述情感来说是重要的字/词对应于语言 B 中的哪些字/词。

步骤 702：根据映射结果产生可以驱动语言 B 的 TTS 的参数，以产生表征情感的语音输出。在此，我们使用语言 B 的情感表述参数表，其根据情感表述参数给出字/词的合成参数。

以上结合具体实施例描述了根据本发明的语音－语音翻译方法。正如本领域一般技术人员所认识别的，本发明还可以用于在同一种语言的不同方言之间进行语音－语音的翻译。如图 8 所示，该过程类似于图 5 所示的翻译过程，区别仅在于，在同种语言不同方言之间进行语音翻译就不再需要文本翻译过程。具体地说包括以下步骤：对一种方言 A 的语音进行识别，形成其文本表示（801）；根据所述文本表示生成另一种方言 B 的语音表示（802）；从方言 A 的语音表示中提取情感表述参数（803）；以及，将情感表述参数检测步骤提取的情感表述参数从方言 A 映射到方言 B；并将映射结果作用于文本－语音生成过程，从而产生可以传达情感的语音输出（804）。

以上结合附图描述了根据本发明优选实施例的具有情感的语音－语音翻译系统和方法。正如本领域技术人员所熟知的，在不背离本发明的精神实质和范围的情况下，本发明可以具有许多修改和变型，本发明将包括所有的这些修改和变型，本发明的保护范围应由所附权利要求书来限定。

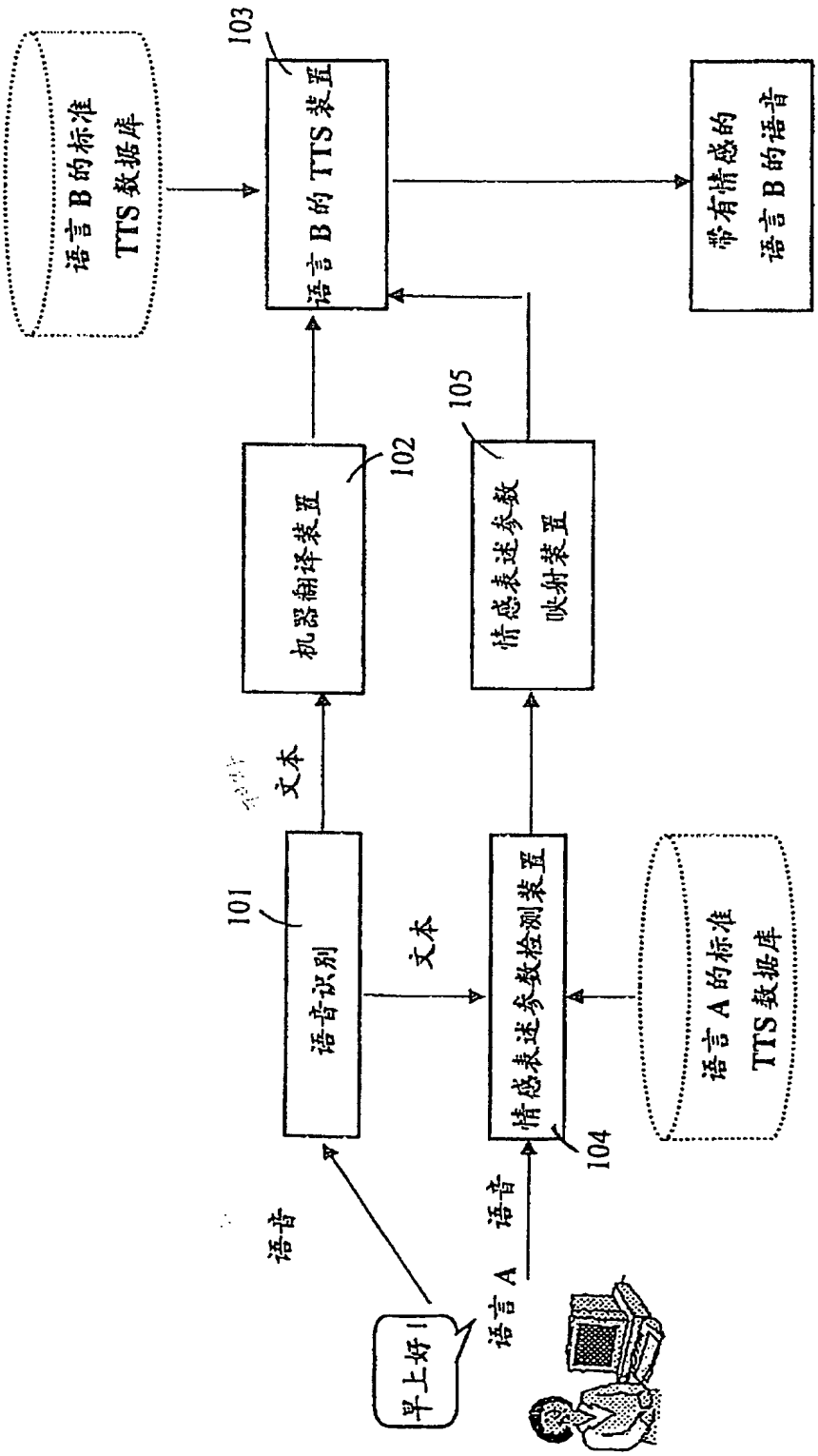

语言B的标准
TTS数据库
103
语言B的TTS装置
带有情感的
语言B的语音
机器翻译装置
102
105
情感表述参数
映射装置
文本
101
语音识别
文本
情感表述参数检测装置
104
语言A的标准
TTS数据库
语音
语音
语言A
早上好！

图1

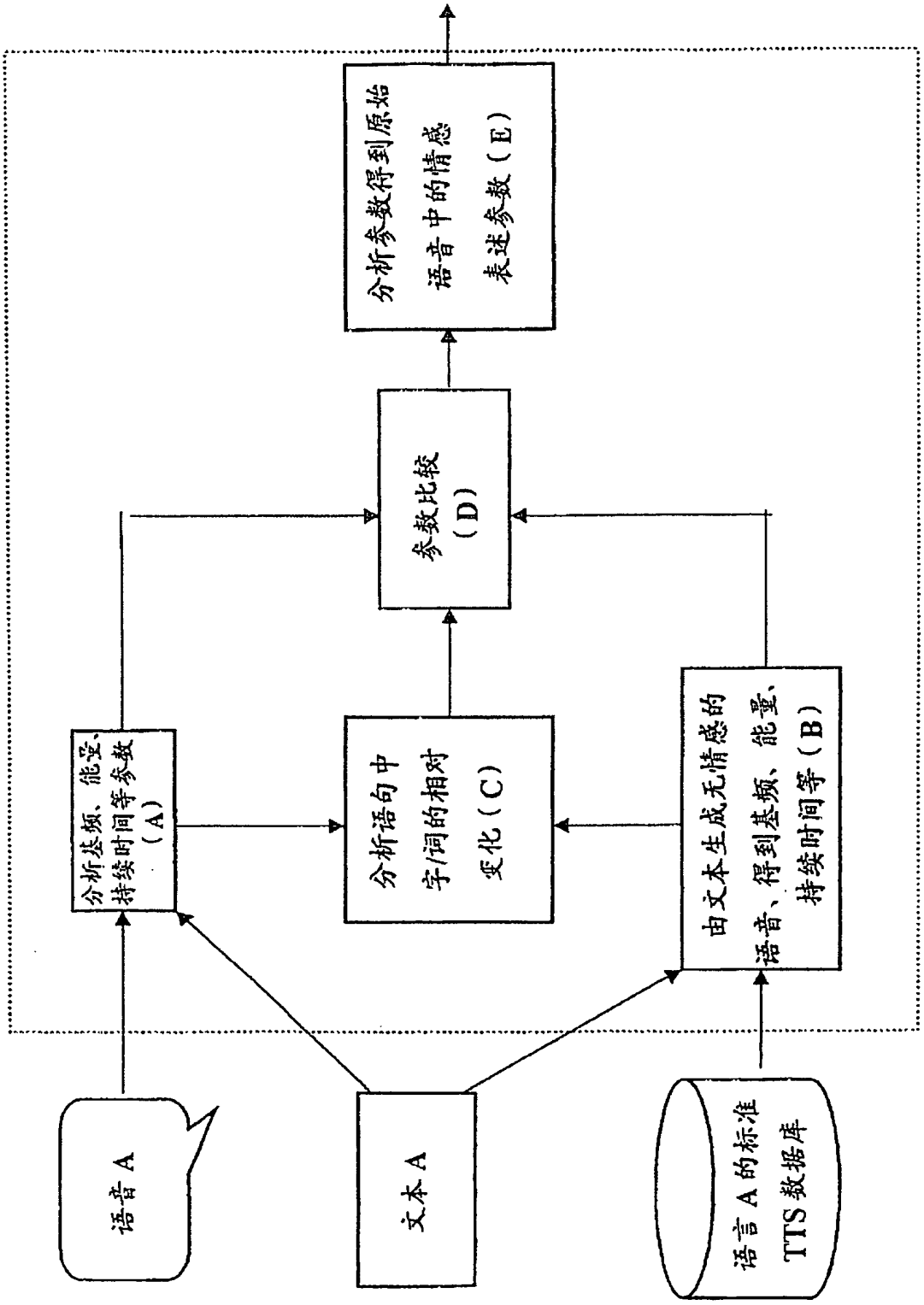

语音 A
文本 A
语言 A 的标准 TTS 数据库
分析基频、能量、持续时间等参数（A）
分析语句中字/词的相对变化（C）
由文本生成无情感的语音、得到基频、能量、持续时间等（B）
参数比较（D）
分析参数得到原始语音中的情感表述参数（E）

图 2

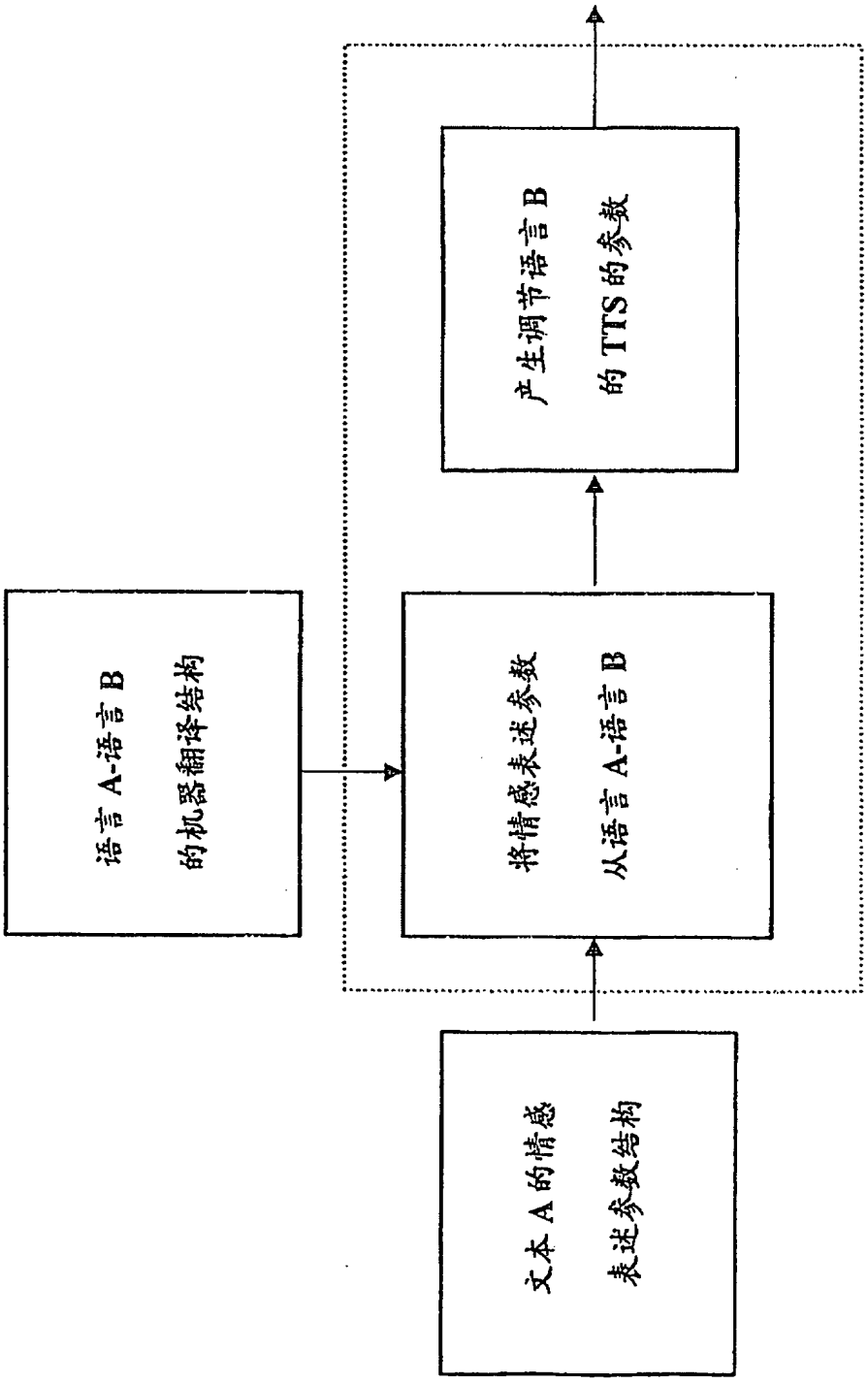

文本 A 的情感
表述参数结构
将情感表述参数
从语言 A-语言 B
语言 A-语言 B
的机器翻译结构
产生调节语言 B
的 TTS 的参数

图 3A

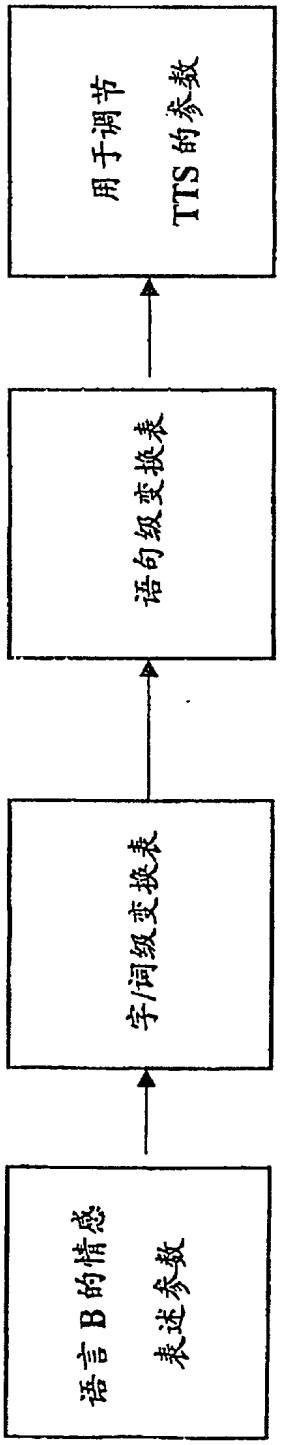


图3B

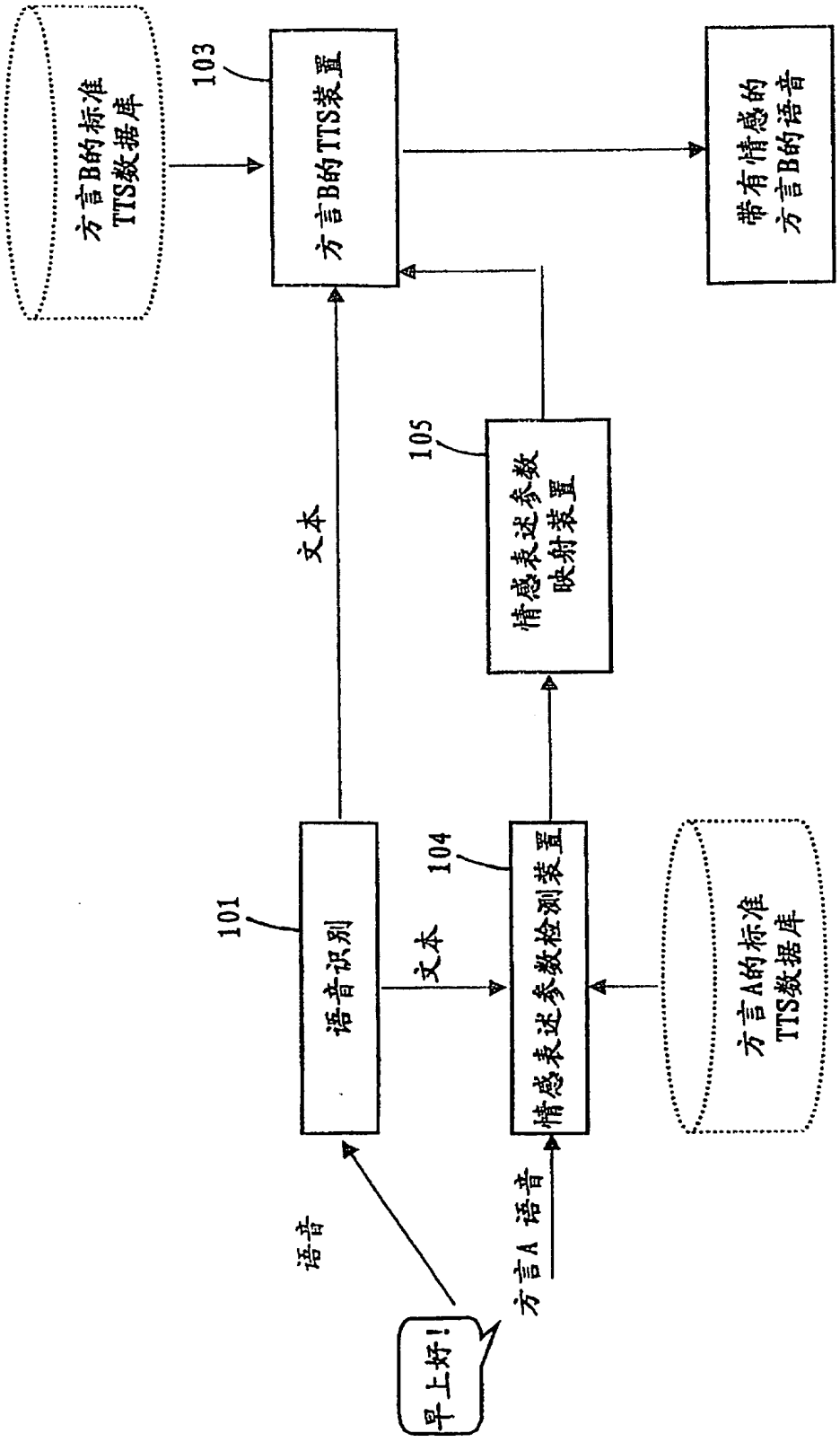

方言B的标准
TTS数据库
103
方言B的TTS装置
带有情感的
方言B的语音
文本
105
情感表述参数
映射装置
101
语音识别
文本
104
情感表述参数检测装置
方言A的标准
TTS数据库
语音
方言A 语音
早上好!

图4

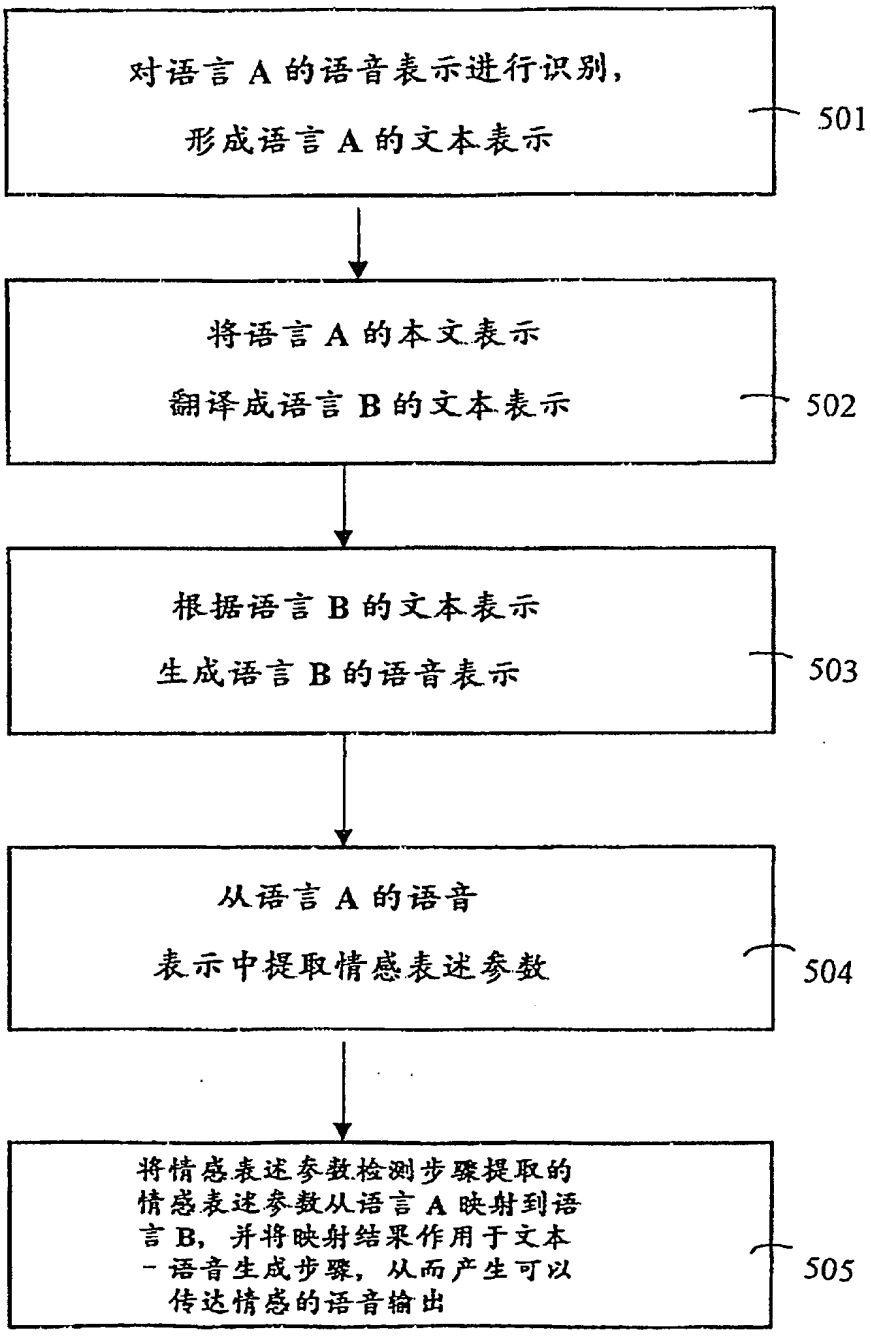

对语言 A 的语音表示进行识别，
形成语言 A 的文本表示
501
将语言 A 的本文表示
翻译成语言 B 的文本表示
502
根据语言 B 的文本表示
生成语言 B 的语音表示
503
从语言 A 的语音
表示中提取情感表述参数
504
将情感表述参数检测步骤提取的
情感表述参数从语言 A 映射到语
言 B，并将映射结果作用于文本
－语音生成步骤，从而产生可以
传达情感的语音输出
505

图5

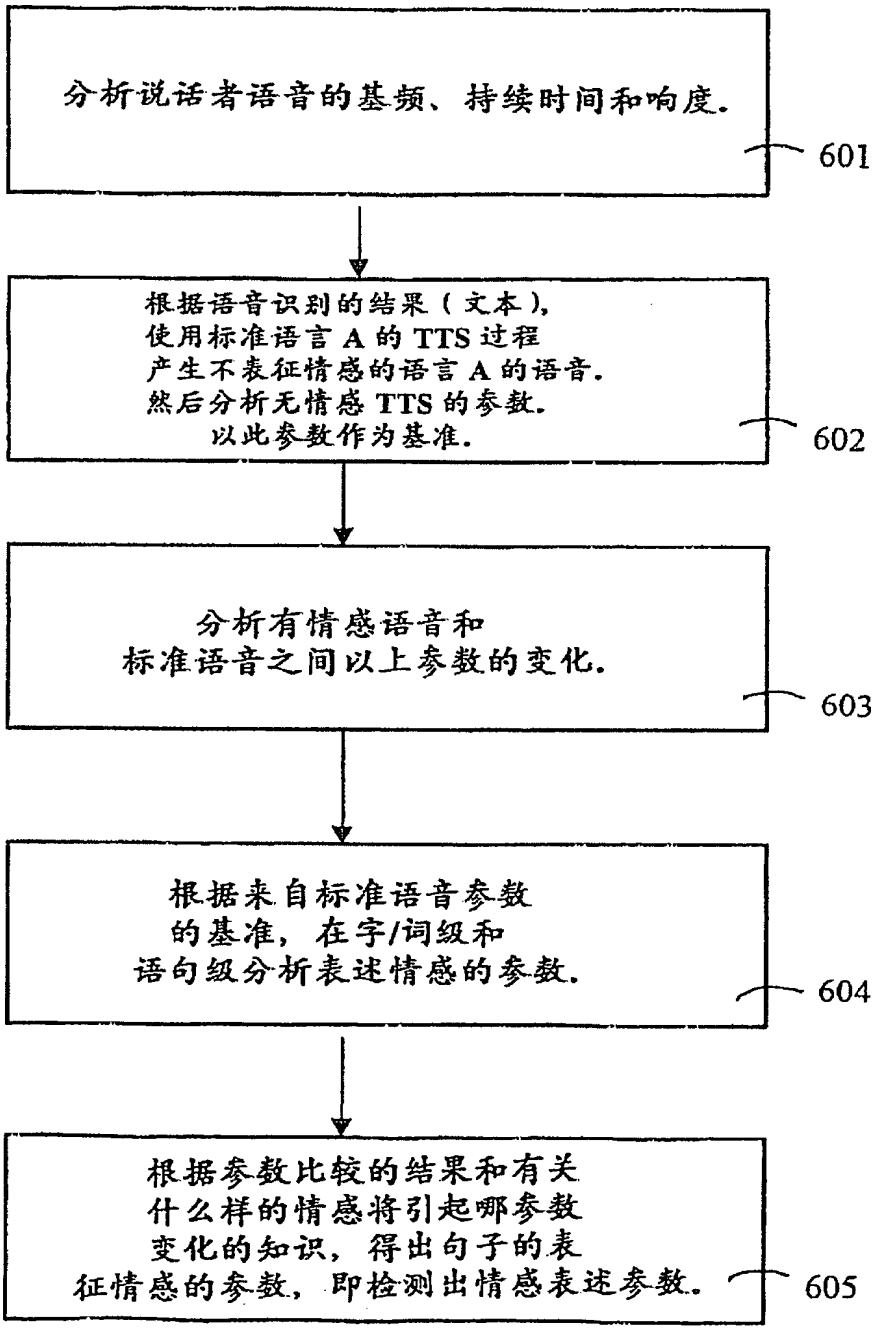

分析说话者语音的基频、持续时间和响度。
601
根据语音识别的结果（文本），
使用标准语言 A 的 TTS 过程
产生不表征情感的语言 A 的语音。
然后分析无情感 TTS 的参数。
以此参数作为基准。
602
分析有情感语音和
标准语音之间以上参数的变化。
603
根据来自标准语音参数
的基准，在字/词级和
语句级分析表述情感的参数。
604
根据参数比较的结果和有关
什么样的情感将引起哪参数
变化的知识，得出句子的表
征情感的参数，即检测出情感表述参数。
605

图6

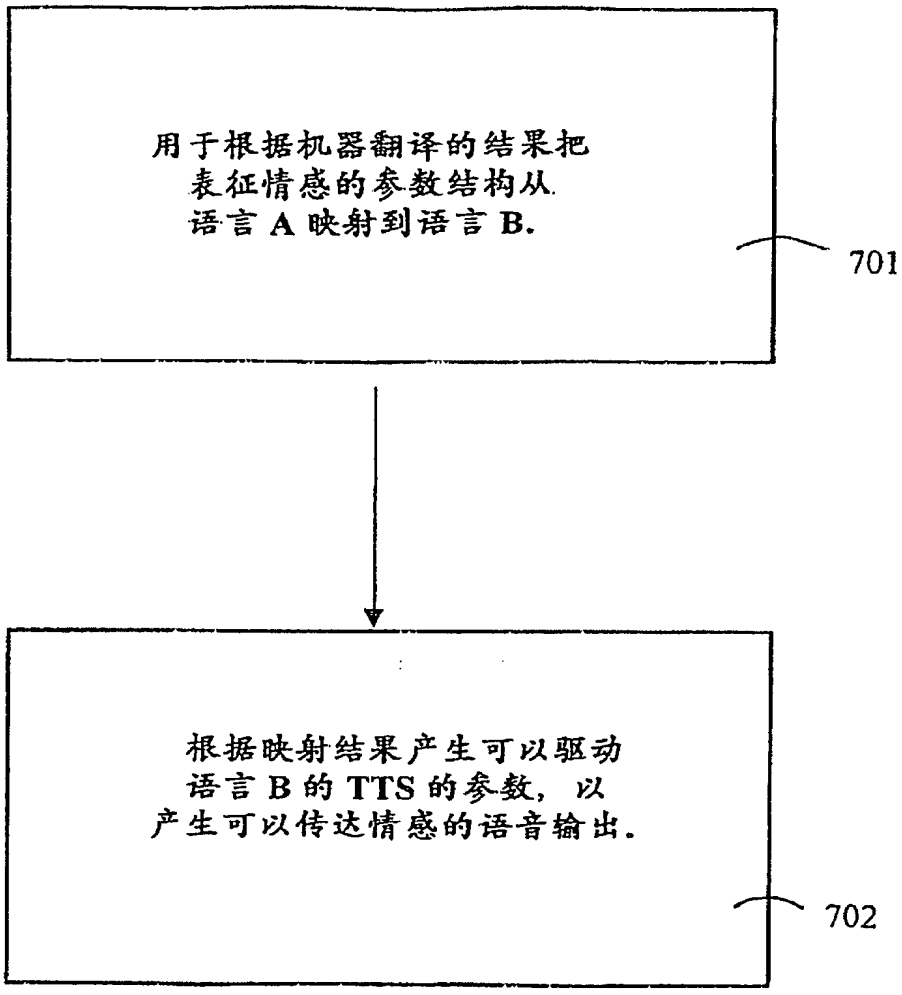

用于根据机器翻译的结果把
表征情感的参数结构从
语言 A 映射到语言 B.
701
根据映射结果产生可以驱动
语言 B 的 TTS 的参数，以
产生可以传达情感的语音输出.
702

图7

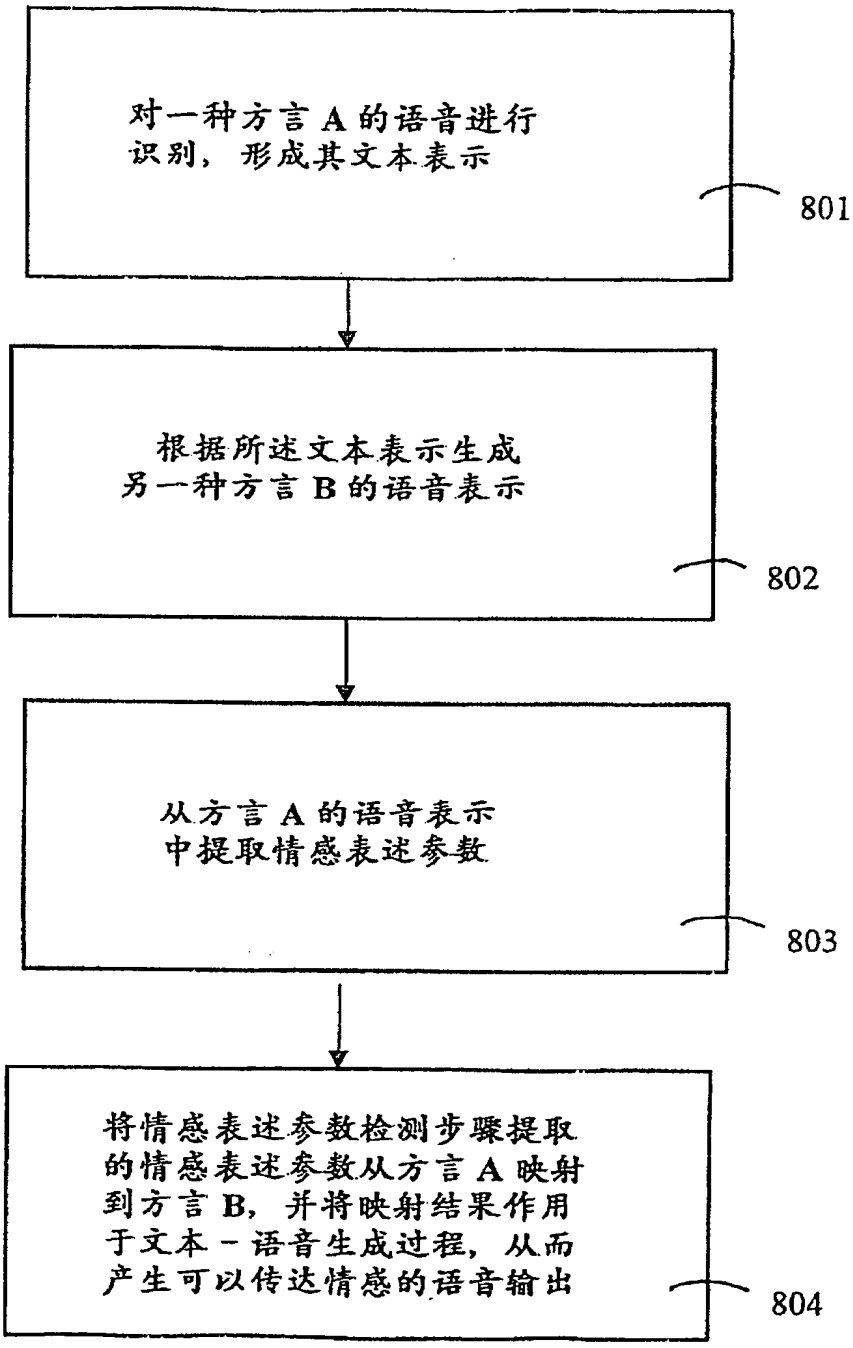

对一种方言A的语音进行识别，形成其文本表示
801
根据所述文本表示生成另一种方言B的语音表示
802
从方言A的语音表示中提取情感表述参数
803
将情感表述参数检测步骤提取的情感表述参数从方言A映射到方言B，并将映射结果作用于文本－语音生成过程，从而产生可以传达情感的语音输出
804

图8